



SCHOOL OF INFORMATION TECHNOLOGY

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA

(University of Technology of Madhya Pradesh)

Mid-Term Examination I June-2023

Discrete Structures (CD201/AL201)

Time: 1:00 Hours

Total Marks: 20

Note: Attempt any 4 questions.

Q.No	Question	Bloom's Taxonomy Level	Marks	COs
Q-1	If $A=\{1,2\}$, $B=\{2,3\}$ and $C=\{3,5\}$, then find $(A \times B) \cap (A \times C)$.	Level(3)	5	CO1
Q-2	A Class has 175 students. The following is the description showing the number of students studying one or more of the following subjects in this class: Mathematics 100, Physics 70, Chemistry 46, Mathematics and Physics 30, Mathematics and Chemistry 28, Physics and Chemistry 23, and none of the three subjects 22 Find b) How many students study all the three subjects?	Level(4)	5	CO1
Q-3	Prove by Mathematical Induction. $7^{2n} + 2^{3n} \cdot 3^{n-1}$ divisible by 25, $\forall n \in \mathbb{N}$ 2^{3n-3}	Level(3)	5	CO1
Q-4	Show that the set of all positive rational numbers forms an abelian group under composition defined by $a * b = \frac{ab}{2}$	Level(3)	5	CO2
Q-5	State and prove that every cyclic group is an abelian group.	Level(2)	5	CO2



SCHOOL OF INFORMATION TECHNOLOGY
RGPV(UTD)
MID SEM I
CD202/CB202 (STATISTICAL METHODS)

TIME: 1:00 HRS

TOTAL MARKS :20

Attempt any 4 questions

Q.no	Questions	Bloom's taxonomy level	Marks	CO																					
Q1	What is Regression analysis? Explain one way and two way ANOVA	1	5	CO 1																					
Q2	Explain Correlation coefficient with properties and example. Find correlation coefficient of given sample data <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td></tr><tr><td>X</td><td>95</td><td>40</td><td>100</td><td>80</td><td>90</td><td>120</td></tr><tr><td>Y</td><td>40</td><td>41</td><td>50</td><td>60</td><td>65</td><td>145</td></tr></table>		A	B	C	D	E	F	X	95	40	100	80	90	120	Y	40	41	50	60	65	145	1	5	CO2
	A	B	C	D	E	F																			
X	95	40	100	80	90	120																			
Y	40	41	50	60	65	145																			
Q3	Write short notes on 1)Type I and Type II errors 2)One tailed and two tailed test 3)Point and Interval Estimation	1	5	CO2																					
Q4	Explain the procedure(steps) for hypothesis testing. What is the significance of p value and how is it calculated?	1	5	CO1																					
Q5	A random sample of 27 pairs of observation gives $r=0.55$. Perform T test and explain how variables in the population are correlated? Given(t at 0.05, df)=2.060	1	5	CO2																					
Q6	Explain ARIMA Model in brief.	1	5	CO4																					



SCHOOL OF INFORMATION TECHNOLOGY

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA
(University of Technology of Madhya Pradesh)

Mid Sem I June-2023

Digital Logic Design (CD203/AL203)

Time: 01 Hour

MM: 20

Note: All questions are compulsory. There is an internal choice among all questions.

Q.No	Questions	Marks	Level
Q-1	Use 2's complement to subtract 1000100 from 1010100. Repeat the same with 1's complement. OR (a) Multiply 1001 by 1101. 1110101 (b) Divide 1110101 by 1001 1101	5	CO1
Q-2	Convert the following numbers from the given base to the bases indicated: decimal 225.225 to binary, octal, and hexadecimal 215.75 OR 327.6 $D7.C$ binary 11010111.110 to decimal, octal, and hexadecimal	5	CO1
Q-3	Simplify the following Boolean functions to a minimum number of literals. Also, specify the theorems or postulates used. (a) $xy + xy'$ x (b) $(x + y)(x + y')$ x OR (a) $xyz + x'y + xyz'$ (b) $zx + zx'y$	5	CO2
Q-4	Simplify the following equation by K-Map (a) $F(A, B, C) = \sum m(0, 2, 4) + \sum d(1, 3, 5, 6, 7)$ (b) Simplify $\pi(0, 2, 3, 6)$ by K-Map OR Minimize the four variable logic function by K-map $F(A, B, C, D) = \sum m(0, 1, 2, 3, 5, 7, 8, 9, 11, 14)$ $\bar{A}D + \bar{A}\bar{B} + \bar{B}D + \bar{B}\bar{C} + ABC\bar{D}$	5	CO2



SCHOOL OF INFORMATION TECHNOLOGY

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA

(University of Technology of Madhya Pradesh)

Mid-Term Examination I June - 2023

Object Oriented Programming (CD-204)

Time: 1:00 Hours

Total Marks:20

Note : All questions are compulsory. Internal choices given with the question.

Q.No	Question	Bloom's Taxonomy Level	Marks	COs
Q-1	Write down the characteristics of Object Oriented Programming. With examples ?	Level(1)	(05)	CO- 1
Q-2	Describe Multiple Inheritance. Write a Program to show multiple inheritance. Also explain diamond problem in multiple inheritance. How this problem can be resolved. OR Describe polymorphism. How it is achieved during compile time and runtime ?	Level(2)	(05)	CO-2
Q-3	Write short note on following (Any 2) 1. Types of Inheritance 2. Constructors and Destructors 3. This pointer 4. Public, protected, private access specifiers. 5. Data Abstraction.	Level(1)	(05)	CO-1 CO-2
Q-4	Explain Inline Function. Also write a program to show utility of Inline Function ?	Level(2)	(05)	CO- 1

About CO column: It stands for Course Outcome. COs are statements that describe what students should be able to do at the end of a course. They are the resultant knowledge skills the student acquires at the end of a course. It defines the cognitive processes a course provides. CO column indicates the Course Outcome number tested through that question. COs are available in the course material provided by faculty.

Course Outcome Statement

CO1	Understand the key features of Object Oriented Programming and Methodology.
CO2	Recognize attributes and methods for given objects.



SCHOOL OF INFORMATION TECHNOLOGY
RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA
(University of Technology of Madhya Pradesh)

Mid-Term Examination-I JUNE-2023
INTRODUCTION TO DATA ANALYTICS CD-205

Time: 1:00 Hours

Total Marks:20

Note : Attempt any **four** Questions.

QNo	Questions	COs	Marks																				
Q-1	The following table shows the distribution of 100 families according to their expenditure per week. The median and mode are given to be Rs. 25 and 24. Calculate the missing frequencies and then arithmetic mean of the data. <table><tr><td>Expend.</td><td>0-10</td><td>10-20</td><td>20-30</td><td>30-40</td><td>40-50</td></tr><tr><td>Families</td><td>14</td><td>?</td><td>27</td><td>?</td><td>15</td></tr></table>	Expend.	0-10	10-20	20-30	30-40	40-50	Families	14	?	27	?	15	CO1	05								
Expend.	0-10	10-20	20-30	30-40	40-50																		
Families	14	?	27	?	15																		
Q-2	Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 27. Assuming that the population is approximately normally distributed with variance 20, can we conclude that the mean is different from 30 years? ($\alpha=0.05$). If the p - value is 0.0340 how can we use it in making a decision?	CO1	05																				
Q-3	Calculate & Fill in the missing entries of the partially completed one way Anova Table. <table><tr><td>Source</td><td>SS</td><td>Df</td><td>MS</td><td>F-statistics</td></tr><tr><td>Between Samples</td><td>2.124</td><td>?</td><td>0.708</td><td>0.75</td></tr><tr><td>With in samples</td><td>?</td><td>20</td><td>?</td><td></td></tr><tr><td>Total</td><td>?</td><td>?</td><td>?</td><td></td></tr></table>	Source	SS	Df	MS	F-statistics	Between Samples	2.124	?	0.708	0.75	With in samples	?	20	?		Total	?	?	?		CO1	05
Source	SS	Df	MS	F-statistics																			
Between Samples	2.124	?	0.708	0.75																			
With in samples	?	20	?																				
Total	?	?	?																				
Q-4	What is data Ecosystem? Explain the components of it.	CO2	05																				
Q-5	Compare data analyst, data engineer and data scientist?	CO2	05																				

$$n=10 \quad \bar{x}=27 \quad \bar{x}=30 \quad \sigma^2=20$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{27 - 30}{\frac{\sqrt{20}}{\sqrt{10}}} = -2.12$$

$$\therefore -1.96 < -2.12$$

Observed value

$$\therefore H_0 \text{ is rejected}$$



SCHOOL OF INFORMATION TECHNOLOGY

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA

(University of Technology of Madhya Pradesh)

Mid-Term Examination II August-2023

Statistical Methods (CD202)

Time: 1:00 Hours

Total Marks: 20

Note: Attempt any 4 questions.

Q.No	Question	Bloom's Taxonomy Level	Marks	COs														
Q-1	The variate x and y are normally correlated with coefficient r and u, v are defined by $u = x\cos\alpha + y\sin\alpha$, $v = y\cos\alpha - x\sin\alpha$ respectively. Show that u and v will be uncorrelated if $\tan 2\alpha = \frac{2r\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}$	Level(2)	5	CO3														
Q-2	Prove that for the binomial population with a density function $f(x, p) = \binom{n}{x} p^x q^{n-x}$, $x = 1, 2, \dots, n$, the maximum likelihood estimator for p is x/n . Find its variance.	Level(2)	5	CO3														
Q-3	Define Kendall's test. Find Kendall's rank correlation coefficient for the given data. <table border="1"><tr><td>x</td><td>2</td><td>7</td><td>1</td><td>5</td><td>8</td><td>10</td></tr><tr><td>y</td><td>4</td><td>5</td><td>6</td><td>8</td><td>10</td><td>9</td></tr></table>	x	2	7	1	5	8	10	y	4	5	6	8	10	9	Level(3)	5	CO4
x	2	7	1	5	8	10												
y	4	5	6	8	10	9												
Q-4	Define time series analysis and forecasting with examples and application.	Level(1)	5	CO4														
Q-5	How to read and write data in R-programming. Give a suitable example.	Level(1)	5	CO5														



SCHOOL OF INFORMATION TECHNOLOGY
RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA
(University of Technology of Madhya Pradesh)

Mid Sem II
Digital Logic Design (CD203/AL203)

Note: All questions are compulsory. There is an internal choice among all questions.

Q.No	Questions	Marks	Level
Q-1	What is the difference between a combinational circuit and a sequential circuit? OR Implement the following function with a multiplexer: $F(A, B, C, D) = \Sigma(0, 1, 3, 4, 8, 9, 15)$.	5	CO3
Q-2	Discuss how JK Master-Slave Flip-Flop works. OR Differentiate between Synchronous and Asynchronous Counters.	5	CO4
Q-3	Define Error-Detection and Error-Correction with examples of its working. OR Differentiate between RTL and DTL?	5	CO5
Q-4	Write short notes on (any two): (a) PAL vs PLA (b) Register (c) Memory	5	CO3,4,5



SCHOOL OF INFORMATION TECHNOLOGY

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA

(University of Technology of Madhya Pradesh)

Mid-Term Examination II July - 2023

Object Oriented Programming (CD-204)

Time: 1:00 Hours

Total Marks:20

Note : All questions are compulsory.

Q.No	Question	Bloom's Taxonomy Level	Marks	COs
Q-1	Write a C++ program to calculate the area of rectangle, square and circle using constructor overloading that accepts input from the user and then shows the desired result	Level(3)	(05)	CO- 3
Q-2	Describe Dynamic memory allocation and de-allocation using suitable example ?	Level(2)	(05)	CO-4
Q-3	Illustrate Stream Class Hierarchy and file handling in C++ using suitable examples ?	Level(1)	(05)	CO-5
Q-4	How exception are handles in C++ programming ? Explain.	Level(2)	(05)	CO- 5



SCHOOL OF INFORMATION TECHNOLOGY

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA

(University of Technology of Madhya Pradesh)

Mid-Term Examination-II August-2023

INTRODUCTION TO DATA ANALYTICS CD-205

Time: 1:00 Hours

Total Marks: 20

Note: Attempt any *four* Questions. Each question carries equal marks.

Q. No	Questions	COs	Marks
Q-1	What is Big data? How big data work?	CO3	05
Q-2	How Big data is Important? Explain 4V's of Big data?	CO3	05
Q-3	Explain Hadoop components in details?	CO4	05
Q-4	Explain Any-5 open source technology of Big data analytics?	CO4	05
Q-5	How to subdividing data perform in hadoop mapreduce? Perform Subdividing process for following sentence using hadoop map reduce? "welcome to rgpv. RGPV is a technical university. There are many departments. SOIT is also a Department."	CO5	05

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

(University of Technology of Govt. of Madhya Pradesh)

B. Tech SECOND SEMESTER (REGULAR/EX) EXAMINATION, AUGUST-2023

Course: Data Science/ AIML

Subject code: CD/AL-201

Subject Title: DISCRETE STRUCTURES

Maximum Marks: 70

Time: 3 hrs.

Note: Answer any FIVE Questions. All Questions carry equal marks.

1.

- a) Let $A = \{2, 3, 4\}$ and $B = \{3, 4, 5, 6, 7\}$. Assume a relation R from A to B such that $(x, y) \in R$ when x divides y . Determine R , its domain and range.
- b) In a survey of 85 people it is found that 31 like to drink milk, 43 like coffee and 39 like tea. Also 13 like both milk and tea, 15 like milk and coffee, 20 like tea and coffee and 12 like none of the three drinks. Find the number of people who like all the three drinks. Display the answer using a Venn diagram.

2.

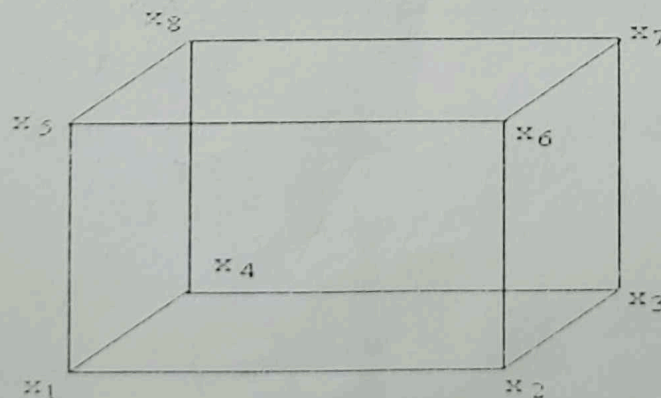
- a) Differentiate between homomorphism and isomorphism of groups with examples.
- b) What is group explain different types of groups

3.

- a) Define tautology and contradiction. Show that "If the sky is cloudy then it will rain and it will not rain", is not a contradiction.
- b) What are the different types of quantifiers? Explain in brief.
Show that $(\exists x)(P(x) \wedge Q(x)) \Rightarrow (\exists x)(P(x) \wedge (\exists x)(Q(x)))$

4.

- a) What is graph? Give an account of the two matrix representations for graphs.
- b) Define a Hamilton path. Determine if the following graph has a Hamilton circuit.



5.

a) Consider the set $A = \{2, 7, 14, 28, 56, 84\}$ and the relation $a \leq b$ if and only if a divides b . Give the Hasse diagram for the poset $(A, \leq)$.

b) Solve the recurrence relation

$$T(k) = 2T(k-1), T(0) = 1$$

6.

a) Let R be the relation on the set of ordered pairs of positive integers such that $((a,b),(c,d)) \in R$ if and only if $ad = bc$. Determine whether R is an equivalence relation or a partial ordering.

b) Test whether 101101, 11111 are accepted by a finite state machine M given as follows : $M = \{Q = \{q_0, q_1, q_2, q_3\}, \Sigma = \{0, 1\}, \delta, q_0 = q_0, F = \{q_0\}\}$
Where δ is

States	Inputs	
	0	1
q_0	q_2	q_1
q_1	q_3	q_0
q_2	q_0	q_3
q_3	q_1	q_2

7.

- Write algorithm to find shortest path in weighted graph.
- Define function explain different types of functions.

8.

Write the short notes (any two)

- Graph Coloring
- Recurrence relation

- Lattice
- Types of functions



TIME: 3:00 Hrs

TOTAL MARKS: 70

Note: Attempt any 5 questions. Each question carries equal marks.

S.No.	Questions	Marks	CO																																
Q-1	a. Show that the line of fit of the following data is given by $y = 0.7x + 11.28$. <table><tr><td>x</td><td>0</td><td>5</td><td>10</td><td>15</td><td>20</td><td>25</td></tr><tr><td>y</td><td>12</td><td>15</td><td>17</td><td>22</td><td>24</td><td>30</td></tr></table> b. Define correlation coefficient. Calculate the coefficient of correlation between the values of x and y. <table><tr><td>x</td><td>78</td><td>89</td><td>97</td><td>69</td><td>59</td><td>79</td><td>68</td><td>61</td></tr><tr><td>y</td><td>125</td><td>137</td><td>156</td><td>112</td><td>107</td><td>136</td><td>123</td><td>108</td></tr></table>	x	0	5	10	15	20	25	y	12	15	17	22	24	30	x	78	89	97	69	59	79	68	61	y	125	137	156	112	107	136	123	108	7+7	CO1
x	0	5	10	15	20	25																													
y	12	15	17	22	24	30																													
x	78	89	97	69	59	79	68	61																											
y	125	137	156	112	107	136	123	108																											
Q-2	a. Explain analysis of variance (ANOVA) one way and two way classification with tables and examples. b. The equation of two regression lines obtained in a correlation analysis in a 60 observations are $5x = 6y + 24$ and $1000y = 768x - 3768$. What is the correlation coefficient and what is the ratio of variances of x and y.	7+7	CO1																																
Q-3	a. State and prove Neyman Pearson Lemma. b. Define maximum likelihood estimator with examples. Write properties of Maximum Likelihood Estimator	7+7	CO2																																
Q-4	a. Define estimation. Explain the properties of good estimators. b. Suppose a random sample of size n is taken from Poisson population with probability function $f(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$. Show that the most powerful critical region of size not exceeding α for testing the hypothesis $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda = \lambda_1$ is of the form $x \leq a$, if $\lambda_0 > \lambda_1$ and $x \geq b$, if $\lambda_0 < \lambda_1$ where $\bar{x}$ is the sample mean and a_0 and b_0 are constants.	7+7	CO2																																



School of Information Technology

Rajiv Gandhi Proudhyogiki Vishwavidyalaya (UTD)

Endsem Examination CD/CB/AL 202 (STATISTICAL METHODS)

S.No.	Questions	Marks	CO																								
Q-5	a. Comparison of parametric and non-parametric inference with examples. Define sign test. b. Define Kendall's test. Find Kendall's rank correlation coefficient for the given data. <table><tr><td>x</td><td>2</td><td>7</td><td>1</td><td>5</td><td>8</td><td>10</td></tr><tr><td>y</td><td>4</td><td>5</td><td>6</td><td>8</td><td>10</td><td>9</td></tr></table>	x	2	7	1	5	8	10	y	4	5	6	8	10	9	7+7	CO3										
x	2	7	1	5	8	10																					
y	4	5	6	8	10	9																					
Q-6	a. A medication is given to 11 patients who suffer from BP issues, their BP are recorded for the analysis. Use Willcoxon signed rank test to test the hypothesis that medication has no effect on BP. <table><tr><td>Before Medication</td><td>112</td><td>113</td><td>118</td><td>120</td><td>119</td><td>113</td><td>110</td><td>122</td><td>126</td><td>115</td><td>119</td></tr><tr><td>After Medication</td><td>116</td><td>120</td><td>117</td><td>125</td><td>126</td><td>111</td><td>111</td><td>117</td><td>126</td><td>112</td><td>129</td></tr></table> b. Define Run test. A foreman for a construction company records injuries reported by workers, during his shift. The following sequence show whether any injury were reported during each month in recent year. 'I' represents injuries which occur at least once in the month and 'N' represent no injury. At $\alpha = 0.05$, can we conclude that the occurrence of injuries in each month is not random. II NNN I NII NNN	Before Medication	112	113	118	120	119	113	110	122	126	115	119	After Medication	116	120	117	125	126	111	111	117	126	112	129	7+7	CO3
Before Medication	112	113	118	120	119	113	110	122	126	115	119																
After Medication	116	120	117	125	126	111	111	117	126	112	129																
Q-7	a. Explain ARIMA model. What is it used for? b. Define time series analysis and forecasting with examples and application	7+7	CO4																								
Q-8	a. Explain any 5 builtin functions in R programming. Define control structures with suitable examples. b. Explain with an example how to read and write data in R program. How lm() function is used to implement linear regression model in R programming?	7+7	CO5																								

**RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA,
BHOPAL**

(University of Technology of Govt. of Madhya Pradesh)

B. Tech SECOND SEMESTER (REGULAR/EX) EXAMINATION, AUGUST-2023

Course: Data Science/ AIML

Subject code: CD/AL-203

Subject Title: DIGITAL LOGIC DESIGN

Maximum Marks: 70

Time: 3 hrs.

Note: Answer any FIVE Questions. All Questions carry equal marks.

Q 1. Describe the number system with suitable example.

Q 2. Design the basic gate with the help of NAND and NOR gate.

XQ 3. Describe the J-K flip flop with the help of S R flip flop and race around condition.

XQ 4. Design the ECL with the performance parameter compare with TTL and Cmos logic.

Q 5. Describe the multiplexers and demultiplexers with suitable diagram.

Q 6. Describe the field programmable gate array.

XQ 7. Design ripple counter with suitable diagram.

Q 8. Define the difference between the PLA and PAL.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

(University of Technology of Govt. of Madhya Pradesh)

B. Tech SECOND SEMESTER (REGULAR/EX) EXAMINATION, AUGUST-2023**Course: Data Science/ AIML****Subject code: CD/AL-204****Subject Title: Object Oriented Programming****Maximum Marks: 70****Time: 3 hrs.****Note:** Answer any **FIVE** Questions. All Questions carry equal marks.

1. a) Differentiate between Object Oriented Programming and Procedure Oriented Programming.
 - b) What is inline function? Write its syntax in C++. When will you make a function inline?
 2. a) Discuss the characteristics of an object oriented programming language.
 - b) Write the purpose of access modifiers? Discuss about the types of access modifiers available in C++.
 - X 3. a) Define constructor. Explain various types of constructors.
 - b) Differentiate between compile time and run time polymorphism.
 4. a) What is meant by function overloading? Give suitable example.
 - b) Write a C++ program to overload the binary operator '+'.
 5. a) What are the advantages of inheritance? Explain various types of inheritance.
 - b) Write the features of friend function in C++. Write a C++ program to demonstrate a friend function.
 6. a) What do you mean by function overriding? Explain with a programming example.
 - b) What are pure virtual functions in C++? Explain with the help of a C++ program.
 - X 7. a) What do you mean by exception? Explain how exception handling is done in C++.
 - b) Explain the various functions which are used to manipulate file pointers.
 - X 8. Write short notes on-
 - a) Stream classes
 - b) New and Delete operators
-

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

(University of Technology of Govt. of Madhya Pradesh)

B. Tech SECOND SEMESTER (REGULAR/EX) EXAMINATION, AUGUST-2023

Course: Data Science

Subject code: CD-205

Subject Title: Object Oriented Programming

Maximum Marks: 70

Time: 3 hrs.

Note: Answer any FIVE Questions. All Questions carry equal marks.

- Q.1 (A) Describe hypothesis testing and list out its basic steps?
- (B) In a factory employing 3000 persons, 5% earn less than Rs. 3 per hour, 580 earn from Rs 3.01 to Rs. 4.50 per hour, 30 percent earn from Rs. 4.51 to Rs. 6.00 per hour, 500 earn from 6.01 to Rs. 7.50 per hour, 20% earn from Rs. 7.51 to Rs. 9.00 per hour and the rest earn Rs. 9.01 or more per hour. What is the median wage? 5.784
- Q.2 (A) Explain Data Ecosystem with its components?
- (B) Describe Roles and responsibilities of data analyst? Compare & Contrast the role of data analyst with data engineer?
- Q.3 (A) Define Big Data & Explain its types?
- (B) Describe 4V's of Big Data? Explain working of big data analytics.
- Q.4 (A) What is Hadoop? Describe any -5 open source technology for big data analytics?
- (B) What is cloud computing? Compare Big data & Cloud computing?
- Q.5 (A) What is disparate data? Explain disparate data sources?
- (B) Describe Map reduce? Explain the components of Map Reduce architecture?
- Q.6 (A) Explain the importance of data extraction? Discuss data extraction techniques?
- (B) Perform Subdividing process for following sentence using Hadoop MapReduce?
"Welcome to RGPV. RGPV is a Technical University. There are many departments. SOIT is also a Department."

- Q.7 (A) Calculate & fill in the missing entries of the partially completed one way ANOVA table.

Source	SS	DF	MS	F-Statistics
Between Samples	2.124	7	0.3034	0.75
Within Samples	7.1888	20	0.3594	
Total	9.3128	27	0.3452	